CS/CNS/EE/IDS 165: Foundations of Machine Learning and Statistical Inference

Introduction to Detection


Anima Anandkumar
Computing and Mathematical Sciences
California Institute of Technology, Pasadena CA 91125
anima@caltech.edu
Copyright ©2013
Outline

Concepts

- Detection and hypothesis testing.
- Deterministic and randomized detectors.
- Cost and risk.
- Bayesian, minimax, and Neyman-Pearson (NP) detectors.

References

The Parameter Space: \( \Lambda \) The parameter \( \theta \) is chosen from one of the \( K \) disjoint subsets

\[
\Lambda = \Lambda_0 \cup \cdots \cup \Lambda_{K-1}, \quad \Lambda_i \cap \Lambda_j = \emptyset
\]

Each \( \Lambda_i \) represents hypothesis \( \mathcal{H}_i \). If \( \Lambda_i \) contains a single element \( \theta_i \), the hypothesis is simple. Otherwise, it is composite. When the hypotheses are composite, the parameter may be random with (conditional) distribution \( \pi_i(\theta) \) within each distribution. If in addition we know the prior probability of each hypothesis, we have

\[
\Theta \sim \sum_{i=0}^{K-1} \Pr(\theta \in \Lambda_i) \pi_i(\theta)
\]

The Statistical Model: \((\Gamma, \mathcal{F}, f(y|\theta))\)

For fixed \( \theta \in \Lambda \), the observation vector \( Y = (Y_1, \ldots, Y_n) \) is distributed according to \( Y \sim f(y|\theta) \).
The Detection (Hypothesis Testing) Problem:
Given realization (data) $Y = y$, detect to which $\Lambda_i$ parameter $\theta$ belongs, or equivalently, test hypotheses $H_i : Y \sim f(y|\theta), \quad \theta \in \Lambda_i$.

Detector: A deterministic detector $\delta(\cdot)$ maps observation $y$ to a hypothesis index in $\{0, \cdots, K-1\}$. It partitions the observation space $\Gamma$ into $K$ disjoint subsets $\Gamma_i$ and identify $\Gamma_i$ with hypothesis $H_i$. When the hypotheses are simple, $\Lambda_i = \{\theta_i\}$, $\delta : \Lambda \to \{\theta_i\}$.

A randomized detector† $\delta(y)$ maps observation $y$ to a probability distribution $\delta(y) = [\delta_1(y), \cdots, \delta_K(y)]$ on $\{0, \cdots, K-1\}$ where $D \sim \delta(y)$ and
$$\delta_i(y) = \Pr(D = i|Y = y)$$

†Deterministic detectors are special cases of randomized detectors
Example: Detecting Signals Under Gaussian Noise

- The possible signals: $\Lambda = \{\theta_1, \theta_2, \theta_3, \theta_4\} \in \mathbb{R}^2$.
- The observation: $Y = \theta + N$, $N \sim \mathcal{N}(0, I)$.
- The likelihood function: $f(y|\theta) = \frac{1}{2\pi} \exp\left\{ -\frac{||y-\theta||^2}{2} \right\}$
- The ML detector is given by

$$\delta(y) = \arg \max_{\theta \in \Lambda} f(y|\theta) = \arg \min_{\theta \in \Lambda} ||\theta - y||$$

An alternative description is the partition of $\mathbb{R}^2$ by

$$\Gamma_i = \{y : ||y - \theta_i|| < ||y - \theta_j||, \forall j \neq i.\}$$

The points on the boundary are assigned arbitrarily.
Example: Detecting the Bias of a Coin

• The Parameter Space: \( \Lambda \)
  \[
  \Lambda = \{ \theta_0 \} \cup \{ \theta_1 \}
  \]
  where \( \theta_0 = \text{“HT”}, \theta_1 = \text{“HH”} \).

• The Observation space \( \Gamma = \{ H, T \} \)

• The statistical model:
  \[
  f(y|\theta_0) = \begin{cases} 
  0.5 & y = H \\
  0.5 & y = T 
  \end{cases}
  \quad f(y|\theta_1) = \begin{cases} 
  1 & y = H \\
  0 & y = T 
  \end{cases}
  \]

• A randomized strategy for detecting bias may be given by the probability \( \delta(y) \) of the detection \( D = 1 \) (HH), i.e.,
  \[
  \delta(y) \triangleq \Pr(D = 1|Y = y)
  \]

Consider, for example,
  \[
  \delta(y) = \begin{cases} 
  0 & y = T \\
  p & y = H 
  \end{cases}
  \]
  where \( p \in [0, 1] \) is the probability that \( \theta_1 \) is chosen when \( y = H \).

• The detection is given by the Bernoulli trial \( D \sim B(\delta(y)) \).
Cost: Each detection \( D = d \) is associated with a cost \( C(d, \theta) \). Since \( D \) is random, the cost \( C_{\delta}(D, \theta) \) is random variable, and it is a function of \( \theta \).

Risk: The risk \( R_{\theta}(\delta) \) of a detector \( \delta \) is the average cost (over the observation and the randomized variables). For a deterministic detector

\[
R_{\theta}(\delta) = \mathbb{E}(C(D, \theta)) = \int f(y|\theta)C(\delta(y), \theta)dy
\]

For a randomized detector, the detector is defined by the PMF \( \delta(y) = [\Pr(D = 0|Y = y), \ldots, \Pr(D = K - 1|Y = y)] \), and the risk is

\[
R_{\theta}(\delta) = \mathbb{E}(C(D, \theta)) = \int f(y|\theta) \sum_{k} \Pr(D = k|Y = y) C_{\delta}(k, \theta)dy
\]

Example: Consider \( \Theta = \{\theta_0, \theta_1\} \). Define the cost by

<table>
<thead>
<tr>
<th>( D = \theta_0 )</th>
<th>( \theta = \theta_0 )</th>
<th>( \theta = \theta_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

The risk when \( \theta = \theta_0 \) is given by

\[
R_{\theta_0}(\delta) = \mathbb{E}(C(D, \theta_0)) = \int_{\{y: \delta(y) \neq 0\}} f(y|\theta_0)dy = \Pr(\delta(Y) \neq \theta_0)
\]

which is the probability of detection error (when \( \theta = \theta_0 \)).
Example: For the bias detection problem, the cost \( C(\hat{\theta}, \theta) \) of detection \( D \sim \delta(y) \) is given by

\[
\begin{array}{|c|cc|}
\hline
& \theta = \theta_0 & \theta = \theta_1 \\
D = \theta_0 & 0 & 100 \\
D = \theta_1 & 100 & 0 \\
\hline
\end{array}
\]

For the randomized detector \( \hat{\theta} \sim B(\delta(y)) \) defined previously,

\[
\begin{align*}
R_{\theta_0}(\delta) &= \Pr(y = T; \theta_0)[C(\theta_0, \theta_0) \Pr(D = \theta_0|y = T) + C(\theta_1, \theta_0) \Pr(D = \theta_1|y = T)] + \Pr(y = H; \theta_0)[C(\theta_0, \theta_0) \Pr(D = \theta_0|y = H) + C(\theta_1, \theta_0) \Pr(D = \theta_1|y = H)] \\
&= 50p \\
R_{\theta_1}(\delta) &= \Pr(y = T; \theta_1)[C(\theta_0, \theta_1) \Pr(D = \theta_0|y = T) + C(\theta_1, \theta_1) \Pr(D = \theta_1|y = T)] + \Pr(y = H; \theta_1)[C(\theta_0, \theta_1) \Pr(D = \theta_0|y = H) + C(\theta_1, \theta_1) \Pr(D = \theta_1|y = H)] \\
&= 100(1 - p)
\end{align*}
\]
Bayesian Risk and Bayesian Detector

The Bayesian formulation assumes a prior distribution of the parameter $\theta \sim \pi(\theta)$. The Bayesian risk is the risk averaged over priors:

$$R(\delta) = \int \pi(\theta) R_{\theta}(\delta) d\theta$$

Note that the Bayesian risk no longer depends on the parameter. The Bayesian detector minimizes the Bayesian Risk

$$\min_{\delta} \mathbb{E}(R_{\theta}(\delta)) = \min_{\delta} \int \pi(\theta) R_{\theta}(\delta) d\theta$$

Example:

Let $\Theta = \{\theta_0, \theta_1\} \sim \pi(\theta)$. Define the cost by

$$C(\delta(y), \theta_i) = \begin{cases} 1 & \delta(y) = i \\ 0 & \text{o.w.} \end{cases}$$

The risk $R_{\theta}(\delta)$ is the conditional error probability

$$R_{\theta}(\delta) = E_{\theta}(C(\delta(Y), \theta)) = \int_{\{y: \delta(y) \neq \theta\}} f(y|\theta) dy = \Pr(\delta(Y) \neq \theta|\Theta = \theta)$$

The Bayesian Risk is the (unconditional) error probability

$$R(\delta) = \pi(\theta_0) \Pr(\delta(Y) \neq \theta_0|\Theta = \theta_0) + \pi(\theta_1) \Pr(\delta(Y) \neq \theta_1|\Theta = \theta_1) = \Pr(\delta(Y) \neq \Theta)$$

Minimizing the Bayesian risk minimizes the detection error probability
The Minimax Detector

**The Criterion**

Minimize the risk for the worst case:

$$\min_{\delta} \max_{\theta \in \Theta} R_\theta(\delta)$$

**Example:**

For the problem of detecting the bias, the minimax detector is given by

$$\min_{\theta(y)} \max \{ R_{\theta_0}(\delta), R_{\theta_1}(\delta) \}$$

This is equivalent to

$$\min_p \max \{ 50p, 100(1-p) \}$$

The optimal $p$ can be computed from

$$50p_{\text{opt}} = 100(1-p_{\text{opt}}) \rightarrow p_{\text{opt}} = \frac{2}{3}$$

The maximum risk is $\$100/3$. 

![Graph showing the relationship between $p$ and $R_{\theta_0}(\delta)$ and $R_{\theta_1}(\delta)$]
The Neyman-Pearson Detector

The Criterion
Minimize the risk for one parameter while imposing constraints on the risks for other parameters.

$$\min_{\delta} R_{\theta K}(\delta) \quad \text{subject to} \quad R_{\theta_i}(\delta) \leq \alpha_i, \ i < K$$

Example
For the problem of flipping a coin with an unknown bias, the Neyman-Pearson detector can be formulated as

$$\min_{\delta} R_{\theta_1}(\delta) \quad \text{subject to} \quad R_{\theta_0}(\delta) \leq 10.$$  

The optimal $\delta$ is given by $p = 1/5$.

A more common formulation is to use probabilities as risks.

$$\min_{\delta} \Pr[D \neq \theta_1; \theta_1] \quad \text{subject to} \quad \Pr[\hat{\theta} \neq \theta_0; \theta_0] \leq \alpha$$
Outline

Main Concepts

- The Bayesian Formulation
- The Bayesian Risk and Conditional Risk.
- Testing Binary Hypothesis.
- Examples

References


The Prior and the Statistical Model

We assume that the random parameter $\Theta$ has a prior distribution $\Theta \sim \pi(\theta)$, which leads to

$$\pi_i \overset{\Delta}{=} \Pr(\Theta \in \Lambda_i) = \int_{\Lambda_i} \pi(\theta) d\theta$$

The statistical model is defined by

$$f(y, \theta) = f(y | \theta) \pi(\theta).$$

Recall the problem statement: Given $Y = y$, test $K$-ary hypothesis

$$\mathcal{H}_i : Y \sim f(y; \theta), \quad \theta \in \Lambda_i, \quad i = 0, \cdots, K - 1$$

The (randomized) detector is defined by

$$\delta(y) = [\delta_k(y)], \quad \delta_k(y) \overset{\Delta}{=} \Pr(D = k | Y = y).$$
Testing based on Priors
Suppose that we don’t take any measurements. Our best bet would be based on the priors

\[ \delta = \arg \max \pi_i, \quad \pi_i = \int_{\Lambda_i} \pi(\theta) d\theta. \]

Testing based on Posterior Distributions
Once we have \( Y = y \), the odds have changed. The posterior distribution of \( \theta \) and posterior probabilities of the hypotheses are given by

\[ f(\theta | y) = \frac{\pi(\theta) f(y | \theta)}{f(y)}, \quad \Pr(\theta \in \Lambda_i | y) = \frac{1}{f(y)} \int_{\Lambda_i} \pi(\theta) f(y | \theta) d\theta \]

Thus the detector that maximizes the probability of correct detection is given by the maximum a posteriori (MAP) detector:

\[ \delta(y) = \arg \max_i \int_{\Lambda_i} \pi(\theta) f(y | \theta) d\theta \]

For \( \Lambda_i = \{\theta_i\} \) with prior \( \pi_i = \Pr(\Theta = \theta_i) \),
\[ \delta(y) = \arg \max_i \pi_i f(y | \theta_i) \]

For simple binary hypotheses, we have the likelihood ratio detector

\[ \delta(y) = \begin{cases} 
1 & \frac{f(y | \theta_1)}{f(y | \theta_0)} > \frac{\pi_0}{\pi_1} \\
0 & \text{otherwise}
\end{cases} \]
Cost Function. Given $\theta$, we define the cost function for each detection $D = i$ by $C(i, \theta)$. The cost is uniform if, for all $\theta \in \Lambda_k$, $C(i, \theta) = C_{ik}$, and the cost function is represented by matrix $C = [C_{ij}]$.

The Risks. For each fixed parameter $\theta \in \Lambda$, the risk of a detector $\delta$ is defined as

$$R_\theta(\delta) = \mathbb{E}(C(D, \theta)|\Theta = \theta) = \sum_k \Pr(D = k)C(k, \theta)$$

$$= \sum_k C(k, \theta) \int \delta_k(y) f(y|\theta) dy$$

The Bayesian risk is the risk averaged over prior $\pi(\theta)$

$$R(\delta) = \mathbb{E}(R_\Theta(\delta)) = \mathbb{E}(C(D, \Theta))$$

$$= \int f(y) \mathbb{E}(C(D, \Theta)|Y = y) dy = \int f(y) R(\delta|y) dy$$

where $R(\delta|y) \overset{\Delta}{=} \mathbb{E}(C(\delta(y), \Theta)|Y = y)$ is the conditional risk, which is a function of $y$.

The Bayesian Detector The Bayesian detector minimizes the Bayesian risk

$$\min_\delta R(\delta) \leftrightarrow \min_\delta R(\delta|y) \leftrightarrow \min_\delta \mathbb{E}(C(D, \Theta)|Y = y)$$

where we note that $D \sim \delta(y)$.
Consider simple binary hypotheses with $\Theta = \{\theta_0, \theta_1\}$ and with priors $\pi_i \triangleq \Pr(\Theta = \theta_i)$. We consider the following cost function

<table>
<thead>
<tr>
<th></th>
<th>$\theta_0$</th>
<th>$\theta_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D = 0$</td>
<td>0</td>
<td>$C_{0,1}$</td>
</tr>
<tr>
<td>$D = 1$</td>
<td>$C_{1,0}$</td>
<td>0</td>
</tr>
</tbody>
</table>

We now argue heuristically the form of the Bayesian detector.

- A sufficient statistic is the likelihood ratio

$$L(y) \triangleq \frac{f(y|\theta_1)}{f(y|\theta_0)}$$

which maps the original n-D statistic to a 1-D statistic.

- It is intuitive that the larger the $L(y)$, the more we should favor $\theta_1$. This suggests a threshold detector on $L(y)$ with the threshold $\tau$ depends on $\{\pi_i, C_{ij}\}$. 

$$L(y) = \frac{f(y|\theta_1)}{f(y|\theta_0)}$$
Some Heuristics

The function for $L(y)$ is given by

$$L(y) = \frac{f(y|\theta_1)}{f(y|\theta_0)}$$

How to choose threshold $\tau$?

- The threshold should be affected by the prior $\pi_i$. If $\pi_0 = \Pr(\Theta = \theta_0)$ is large, the region for detecting $\theta_0$ should be larger, i.e., $\tau$ larger. This should be made relative to $\pi_1$. One possible choice is

$$\tau = \frac{\pi_0}{\pi_1}$$

In particular, such a choice makes sense when $\pi_0 = 0$.

- The threshold should also be affected by costs; if $C_{01}$ is large, then the cost of mistakenly detecting $\theta_1$ is high. Thus the region for detecting $\theta_1$ should be reduced, and $\tau$ should be smaller. This again needs to be made relative to the cost $C_{10}$. One possible choice, incorporating the effect of priors, is

$$\tau = \frac{\pi_0 C_{10}}{\pi_1 C_{01}}$$

The choice above makes sense, but it gives no indication that it is in any way optimal.
**Theorem**

The Bayesian risk is minimized by a deterministic detector

\[
\delta_k(y) = \begin{cases} 
1 & k = k_o \\
0 & \text{o.w.} 
\end{cases}
\]

(1)

where \(k_o\) minimizes the conditional cost

\[
k_o = \arg\min_k \mathbb{E}(C(k, \Theta) | Y = y)
\]

\[
= \arg\min_k \int C(k, \theta) f(y | \theta) \pi(\theta) d\theta
\]

**Remark:**

The Bayesian detector minimizes the cost averaged over parameter \(\Theta\) and conditioned on data \(y\).

Proof: We fix \(y\) and we need to find the optimal

\[
\delta_k(y) \triangleq \Pr(D = k | Y = y), \quad k = 0, \cdots, K - 1
\]

that minimizes

\[
R(\delta|y) \triangleq \mathbb{E}(C(D, \Theta) | Y = y)
\]

\[
= \mathbb{E}(\mathbb{E}(C(D, \Theta) | D, Y = y))
\]

\[
= \sum_k \Pr(D = k | Y = y) \mathbb{E}(C(k, \Theta) | D = k, Y = y))
\]

\[
= \sum_k \delta_k(y) \mathbb{E}(C(k, \Theta) | Y = y)) = \sum_k \delta_k(y) c_k(y)
\]

where the last equality comes from the fact that \(\Theta \rightarrow Y \rightarrow D\) forms a Markov chain, i.e., conditioned on \(Y\), \(\Theta\) and \(D\) are independent.
We can formulate this optimization as a linear programming problem:

\[
\begin{align*}
\text{minimize} & \quad \sum_{k=0}^{K-1} \delta_k(y) c_k(y) \\
\text{subject to} & \quad \sum_{k=0}^{K-1} \delta_k(y) = 1, \quad \delta_k(y) \geq 0, \quad \forall k
\end{align*}
\]

From the Fundamental Theorem of Linear Programming\(^\dagger\), The solution to this optimization is given in (1).

\[
\sum_k c_k(y) \delta_k(y) = \gamma
\]

The linear programming problem can be illustrated geometrically. The minimization corresponds to moving the plane defined by 
\[
\sum_k c_k(y) \delta_k(y) = \gamma
\]
up from \(\gamma = 0\) until it touches first the the polyhedron defined by \(\sum_k \delta_k(y) = 1\). The optimal \(\delta_k(y)\) must be one of the extreme points of the polyhedron.

For simple hypothesis $\Lambda = \{\theta_0, \theta_1\}$ with priors $\pi_i = \Pr(\theta = \theta_i)$ and cost function $C_{ij} \triangleq C(i, \theta_j)$, the conditional costs are given by

$$
\mathbb{E}(C(1, \theta)|Y = y) = \frac{1}{f(y)}(C_{10}f(y|\theta_0)\pi_0 + C_{11}f(y|\theta_1)\pi_1)
$$

$$
\mathbb{E}(C(0, \theta)|Y = y) = \frac{1}{f(y)}(C_{00}f(y|\theta_0)\pi_0 + C_{01}f(y|\theta_1)\pi_1)
$$

Hence $\delta(y) = 1$ if

$$
f(y|\theta_1)\pi_1(C_{01} - C_{11}) \geq f(y|\theta_0)\pi_0(C_{10} - C_{00})
$$

Without loss of generality, assume $C_{ij} > C_{ii}$ for all $i, j$. We have

$$
\delta(y) = 1 \text{ if } \frac{f(y|\theta_1)}{f(y|\theta_0)} \geq \frac{\pi_0(C_{10} - C_{00})}{\pi_1(C_{01} - C_{11})} \triangleq \tau_B
$$

Remarks

- The detector is given by the likelihood ratio (sufficient statistic for $\theta$).
- If $\theta_0$ is more likely, $\pi(\theta_0)$ is bigger and the threshold for detecting $\theta_1$ is higher.
- When $C_{ii} = 0$, bigger the $C_{10}$, higher the threshold of detecting $\theta_1$.
- When the two hypothesis are equally likely, and $C_{i1} = C_{i0}$, $\tau_B = 1$, which implies that the detector picks the hypothesis that is more likely.
Simple Binary Hypothesis: The Risk

The Decision Region
The optimal detector partitions the observation space into \( \Gamma_0 \) and \( \Gamma_1 \) according to

\[
\Gamma_1 \triangleq \{ y : f(y|\theta_1) \geq \tau_B f(y|\theta_0) \}, \Gamma_0 = \Gamma_1^c
\]

The Bayesian Risk
The Bayesian risk of the optimal detector is given by

\[
R(\delta) = \pi_0[C_{00} \int_{\Gamma_0} f(y|\theta_0)dy + C_{10} \int_{\Gamma_1} f(y|\theta_0)dy] + \\
\pi_1[C_{01} \int_{\Gamma_0} f(y|\theta_1)dy + C_{11} \int_{\Gamma_0} f(y|\theta_1)dy]
\]

A Special Cost
For the 0-1 cost, i.e., \( C_{ij} = 1 \) when \( i \neq j \) and \( C_{ii} = 0 \), we have

\[
R(\delta) = \pi_0 \int_{\Gamma_1} f(y|\theta_0)dy + \pi_1 \int_{\Gamma_0} f(y|\theta_1)dy \\
= \Pr(\text{detection error})
\]
Example: Detection of 2 Signals

Signal in Gaussian Noise
Consider the transmission of one of the two signal $\theta_0$ and $\theta_1$ over a Gaussian channel

$$Y = \Theta + N, \Theta \in \{\theta_0, \theta_1\}, N \sim \mathcal{N}(0, \sigma^2).$$

The prior probability is given by $\pi_i = \Pr(\Theta = \theta_i)$.

The simple binary hypothesis can be expressed as

- $\mathcal{H}_0 : Y \sim \mathcal{N}(\theta_0, \sigma^2)$,
- $\mathcal{H}_1 : Y \sim \mathcal{N}(\theta_1, \sigma^2)$.

Cost and Risk
Suppose that we have the cost, i.e.,

$$C_{ij} = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases}$$

The Bayesian risk for this cost is the error probability.

The Likelihood Functions

$$f(y|\theta_i) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{(y - \theta_i)^2}{2\sigma^2}\right\}.$$
The observation space is partitioned into $\Gamma_0$ and $\Gamma_1$ where

$$\Gamma_1 \triangleq \{ y : \frac{f(y|\theta_1)}{f(y|\theta_0)} \geq \frac{\pi_0}{\pi_1} \}.$$  

$$= \{ y : \ln f(y|\theta_1) - \ln f(y|\theta_0) \geq \ln \frac{\pi_0}{\pi_1} \}$$

$$= \{ y : y \geq \theta_0 + \frac{\theta_1}{2} + \frac{\sigma^2}{\theta_1 - \theta_0} \ln \frac{\pi_0}{\pi_1} \triangleq \gamma \}$$

Remarks

• When $\pi_0 = \pi_1$, the decision boundary $\gamma$ lies between the means of the two density functions. The boundary moves with $\pi_i$ to favor the more likely hypothesis.

• For really noisy data, $\sigma \to \infty$, the detection goes with the prior. On the other hand, for really clean data, the detection ignores the prior.
The Baysian risk (the error probability) is given by

$$R(\delta) = \pi_0 Q(\frac{\gamma - \theta_0}{\sigma}) + \pi_1 Q(\frac{\theta_1 - \gamma}{\sigma})$$

where the $Q$-function is the tail probability

$$Q(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

**Remarks**

The error probability does not depend on the actual locations of $\theta_i$; it depends on their relative distance $\theta_1 - \theta_0$:

$$\frac{\gamma - \theta_0}{\sigma} = \frac{\theta_1 + \theta_0}{2\sigma} + \frac{\sigma}{\theta_1 - \theta_0} \ln \frac{\pi_0}{\pi_1}$$
$$\frac{\theta_1 - \gamma}{\sigma} = \frac{\theta_1 + \theta_0}{2\sigma} - \frac{\sigma}{\theta_1 - \theta_0} \ln \frac{\pi_0}{\pi_1}$$

When the two hypotheses are equally likely, we have

$$\gamma = \frac{\theta_1 + \theta_0}{2}, \quad R(\delta) = Q\left(\frac{\theta_1 - \theta_0}{2\sigma}\right)$$
Consider the following binary hypothesis:

\[ H_0 : \ Y = \Theta + N, \Theta \in \Lambda_0 = \{\theta_{00}, \theta_{01}\} \]

\[ H_1 : \ y = \Theta + N, \Theta \in \Lambda_1 = \{\theta_{10}, \theta_{11}\} \]

where \( N \sim \mathcal{N}(0, \sigma^2) \). Suppose also that \( \pi_{ij} \overset{\Delta}{=} \Pr(\Theta = \theta_{ij}) = \frac{1}{4} \), and the cost is uniform, i.e.,

\[
C_{i,\theta} = \begin{cases} 0 & \theta \in \Lambda_i, \\ 1 & \text{otherwise} \end{cases}
\]

The optimal detector is given by

\[
\delta(y) = \begin{cases} 0 & \sum_{i,j} C(0, \theta_{ij}) f(y|\theta_{ij}) \pi_{ij} < \sum_{i,j} C(1, \theta_{ij}) f(y|\theta_{ij}) \pi_{ij} \\ 1 & \sum_{i,j} C(1, \theta_{ij}) f(y|\theta_{ij}) \pi_{ij} < \sum_{i,j} C(0, \theta_{ij}) f(y|\theta_{ij}) \pi_{ij} \end{cases}
\]

Substituting \( \pi_{ij} \) and \( C(i, \theta) \), the detector is given by the test the ratio

\[
L(y) = \frac{\pi_{10} f(y|\theta_{10}) + \pi_{11} f(y|\theta_{11})}{\pi_{00} f(y|\theta_{00}) + \pi_{01} f(y|\theta_{01})} = \frac{e^{-\frac{(y-\theta_{10})^2}{2\sigma^2}} + e^{-\frac{(y-\theta_{11})^2}{2\sigma^2}}}{e^{-\frac{(y-\theta_{00})^2}{2\sigma^2}} + e^{-\frac{(y-\theta_{01})^2}{2\sigma^2}}}
\]

Suppose that \( \theta_{00} = -\theta_{11} \) and \( \theta_{01} = -\theta_{10} \). Intuition suggests that

\[
\Gamma_0 = \{y : y < 0\}, \quad \Gamma_1 = \{y : y \geq 0\},
\]

and it is true.
Recall the coin tossing problem $Y \sim p(y; \theta)$ with binary hypotheses

$$
\mathcal{H}_0 : f(y|\theta_0) = \begin{cases} 
0.5 & y = H \\
0.5 & y = T
\end{cases} \quad \text{vs.} \quad \mathcal{H}_1 : p(y; \theta_1) = \begin{cases} 
1 & y = H \\
0 & y = T
\end{cases}
$$

The cost is given by $C_{ii} \triangleq C(i, \theta_i) = 0$ and $C_{ij} \triangleq C(i, \theta_j) = 100$ for $j \neq i$. The Bayesian detector is based on testing the likelihood ratio against the threshold

$$
\tau_B = \frac{\pi_0(C_{10} - C_{00})}{\pi_1(C_{01} - C_{11})} = \frac{\pi_0}{\pi_1}.
$$

Because we have discrete random variables, we need to do this test individually for $y = T$ and $y = H$.

$$
y = T \quad \frac{f(T|\theta_1)}{f(T|\theta_0)} = 0 \leq \tau_B \rightarrow \delta(T) = 0.
$$

$$
y = H \quad \frac{f(H|\theta_1)}{f(H|\theta_0)} = 2 > \tau_B
$$

As expected, the Bayesian detector depends on the prior $\pi_0$.

$$
\delta_B(T) = 0, \quad \delta_B(H) = \begin{cases} 
0 \quad \pi_0 \geq \frac{2}{3} \\
1 \quad \pi_0 < \frac{2}{3}
\end{cases}
$$

The optimal Bayesian detector is given by

$$
R(\delta_B) = E\{\delta(Y)C_{1,\theta} + (1 - \delta(Y))C_{0,\theta}\}
$$

$$
= \pi_0 P(\delta_B = 1|\theta_0) \times 100 + \pi_1 P(\delta_B = 0|\theta_1) \times 100
$$

$$
= \begin{cases} 
100\pi_1 \quad \pi_0 \geq \frac{2}{3} \\
50\pi_0 \quad \pi_0 < \frac{2}{3}
\end{cases} \rightarrow \max_{\pi_0} R(\delta_B) = \frac{100}{3}
$$
**Summary**

**Assumption:** $\Theta \sim \pi(\theta)$

**Key Steps:** Minimizing the conditional cost:

$$d = \arg \min_i \mathbb{E}(C(i, \theta)|y)$$

where the average is taken over $\Theta$ using conditional distribution $f(\theta|y)$. For simple binary hypothesis

$$\frac{f(y|\theta_1)}{f(y|\theta_0)} \geq \frac{\pi_0(C_{10} - C_{00})}{\pi_1(C_{01} - C_{11})}$$

**Pros and Cons:**

- The knowledge of $\pi(\theta)$ allows a simple tradeoff among risks $R_\theta(\delta)$.

- Bayesian detectors are widely used in digital communications where one can design the prior of the hypotheses (messages).

- Due to their simplicity, Bayesian detectors often serve as a mathematical device to obtain other results when the prior is unknown.

- The prior may not be available in practice, which limits the application of Bayesian detectors.

- If the prior is not accurate, the performance is not robust.
Problem Formulation

\( \Lambda, \Lambda_i \) Parameter space. \( \Lambda = \bigcup \Lambda_i \)

\( \pi(\theta), \pi_i \) Prior probability density or mass function. \( \pi_i = \Pr(\Theta \in \Lambda_i) \)

\( \mathcal{H}_i \) Hypothesis \( i \).

\( f(y|\theta) \) Conditional PDF/PMF. Likelihood function.

\( f(y|\mathcal{H}_i), f_i(y) \) Conditional PDF/PMF. Likelihood function under \( \mathcal{H}_i \)

\( L(y; \theta_i, \theta_j) \) Likelihood ratio \( L(y; \theta_i, \theta_j) = \frac{f(y|\theta_i)}{f(y|\theta_j)} \)

\( l(y; \theta_i, \theta_j) \) Log-likelihood ratio \( l(y; \theta_i, \theta_j) = \log L(y; \theta_i, \theta_j) \)

\( \Gamma \) Observation space.

\( \Gamma_i \) Decision region for \( \mathcal{H}_i \).

Detector

\( \delta(y) \) Decision function (probability vector) that maps \( y \) to \( \Pr(D = i|Y = y) \).

For deterministic detectors, \( \delta(y) \) maps \( y \) directly to \( \mathcal{H}_i \).

\( \delta_{B,\pi} \) the Bayesian detector for prior \( \pi \).

 Costs and Risks

\( C(i, \theta) \) Cost of detection \( i \) associated with parameter \( \theta \).

\( C_{ij} \) Uniform cost: \( C_{ij} = C(i, \theta), \theta \in \Lambda_j \).

\( r(y) \) Likelihood ratio: \( l(y) = \frac{p(y;\theta_i)}{p(y;\theta_j)} \)

\( R_{\theta}(\delta) \) Risk associated with parameter \( \theta \): \( R_{\theta}(\delta) = E\{\delta(y)C(1, \theta) + (1 - \delta(y))C(0, \theta)\} \)

\( R(\delta) \) Bayesian risk of \( \delta \): \( R(\delta) = \int p(\theta)R_{\theta}(\delta) \).

\( R(\delta|y) \) Conditional Bayesian risk: \( R(\delta|y) = E\{\delta(y)C(1, \theta) + (1 - \delta(y))C(0, \theta)|y\} \)

\( V(\pi_0) \) The minimum bayesian risk: \( V(\pi_0) = \pi_0 R_{\theta_0}(\delta_{B,\pi_0}) + (1 - \pi_0) R_{\theta_1}(\delta_{B,\pi_0}) \)